Hidden Markov Models for Wind Farm Power Output

D. Bhaumik*, D. Crommelin†, S. Kapodistria‡ Member, IEEE, and B. Zwart§ Member, IEEE,

Abstract—The reliability of the transmission grid is challenged by the integration of intermittent renewable energy sources into the grid. For model-based reliability studies, it is important to have suitable models available of renewable energy sources like wind and solar power. In this study, we investigate to what extent the power output of wind farms can be modeled with discrete Hidden Markov Models (HMMs).

The parameters of the HMMs are inferred from measurement data from multiple turbines in a wind farm. We use these models both for individual turbine output and for total aggregated power output of multiple turbines. When modeling individual turbine output, the hidden process in the HMM is instrumental in capturing the dependencies between the output of the different turbines. It is important to account for these dependencies in order to correctly capture the upper quantiles (90%, 95%, 99%) of the distribution of the wind farm aggregated power output. We show that despite their simple structure, HMMs are able to reproduce important features of the power output of wind farms. This opens up possibilities to model and analyze these features with methods and techniques stemming from the field of Markov models and stochastic processes.

Index Terms—Wind farm power generation, data models, power system modeling, time series analysis, hidden Markov models.

*D. Bhaumik is a PhD student in the Scientific Computing group at Centrum Wiskunde and Informatica (CWI), Science park 123, Amsterdam, 1098 XG, Netherlands (email: bhaumik@cwi.nl).
†D. Crommelin is a Professor at the Korteweg-de Vries Institute for Mathematics of the University of Amsterdam and leads the Scientific Computing group at CWI Amsterdam (email: daan.crommelin@cwi.nl.)
‡S. Kapodistria is an assistant professor in the Stochastics group of the Department of Mathematics and Computer Science in Eindhoven University of Technology (email: s.kapodistria@tue.nl). The work of Stella Kapodistria is supported by the NWO Gravitation Program Networks and the TKI-WoZ DAISY4Offshore project of the Dutch government.
§B. Zwart is a Professor at the Department of Mathematics and Computer Science in Eindhoven University of Technology and is the group leader of the Stochastics group at CWI Amsterdam (email: Bert.Zwart@cwi.nl).

Nomenclature

P_t^w	t-th datum of power output of the w-th turbine, $t = 1, \ldots, T$.
G_t	t-th datum of total power produced, $t = 1, \ldots, T$.
R_P	Rated power of a wind turbine.
G^*	Quantile threshold.
γ_P	Quantile fraction for measurement data, i.e., $\gamma_P := \frac{1}{T} \sum_{t=1}^{T} \mathbb{1}(G_t > G^*)$.
Y_t^w	Stochastic output process for the power output of the w-th turbine at time t, $t = 1, 2, \ldots$.
X_t	Hidden process at time t, $t = 1, 2, \ldots$.
p	Number of free parameters in a model.
$\mathcal{L}$	Likelihood function.

I. Introduction

THE necessity to reduce carbon footprints has led to a need for integrating renewable energy sources like wind turbines and photo-voltaic arrays into the power grid. These renewable energy sources, being unpredictable in nature, inject power intermittently into the grid, thereby challenging the reliability of the power network, cf. [1]–[3]. It is therefore of great importance, for both the planning and the operation of grids, to develop appropriate stochastic models of power systems with intermittent, renewable power generation. This is particularly important in reliability studies [4], [24].

In this study, we focus on the occurrence of high levels of wind farm power output. We purposefully chose wind farms as they are the largest source of intermittent renewable energy worldwide. Furthermore, assessing the upper quantiles of their power output is relevant for grid reliability assessment [5]. This topic has received considerable attention and various approaches have been proposed in the literature. When utilizing stochastic models to capture wind power generation, two choices must be made: (i) whether to model the power output directly or model the wind speed and employ a separate model

(e.g., a wind power curve (WPC)) to obtain the power output from the wind speed; and (ii) whether to model the aggregated wind farm output or model the individual power output from the wind turbines in the farm.

Regarding the first choice, there is a significant body of literature, see, e.g., [8]–[14] and the references therein, that use some statistical or stochastic method to model the wind and map the wind speed to the power output using the WPC. Such indirect models do not effectively capture the stochasticity of the power output for a given wind speed [15]. The direct modeling of power output is considered in e.g., [16]. In that study, a limited autoregressive integrated moving average (LARIMA) model is used for the modeling of the individual wind turbine power, and a multivariate-ARIMA model is used for the modeling of the cross-correlations between the produced power for different parts of the wind farm. Furthermore, [17] and [18] use discrete-time Markov chains (DTMCs) for the modeling of the power output of an individual turbine.

Regarding the second choice, in order to model the joint or the aggregated power output of a wind farm, it is paramount to consider some stochastic dependence structure between the power output of the turbines, see, e.g., [19]–[21]. These approaches, are oftentimes too complicated to be used for simulations, or for theoretical studies concerning the design and the planning of power grids. Another interesting approach that can be deployed for the short-term forecasting of the joint or the aggregated power output is that of deep neural networks (DNNs). In [22], DNNs are used to model wind speed and, in [23], DNNs based on meta-regression are used for wind power forecasting. DNNs require a trove of available data for training and a concrete understanding of the system at hand so as to carefully choose the type of the neural network. Moreover, DNNs have a very complicated structure which is not suitable for theoretical analysis. This was also reported in [6], in which the authors discuss the use of DTMCs to model the wind speed.

For the above mentioned reasons, we propose, in this study, a class of stochastic models called Hidden Markov Models (HMMs) that are able to accurately capture the distribution of the power output, and are also simple enough to be amenable to analysis and to Monte Carlo simulation. In particular, the theoretical methods and techniques from the field of Markov models and stochastic processes (oftentimes appearing in the study of communication and transportation networks) provide a rich set of tools that can be employed for analyzing power systems, [7]. Such simple stochastic models are particularly useful for model-based risk and reliability studies.

To assess the accuracy of HMM's in fitting the upper quantiles ($90\%, 95\%, 99\%$) of the observed power output of a wind farm, we use time series power output measurements from six individual turbines from an existing on-shore wind farm. One advantage of inferring models directly from the measured power output is that they then reflect various factors that have an impact on the power generation: not only the variability of the wind speed, but also e.g., curtailing of wind turbines, shut-down of turbines for maintenance or for other operational purposes, wear and tear of the equipment, etc.

We consider HMMs for both individual turbine outputs as well as aggregated power outputs. We refer to the former as the microscopic approach, in which we model the individual wind turbine outputs, while aiming to account for dependencies of the output between different turbines. By appropriately aggregating their outputs, we study the aggregated power output and its upper quantiles. This approach is flexible, as it is easy to implement when considering the impact of adding more turbines to a farm or shutting down turbines (e.g., for maintenance). By contrast, with the macroscopic approach, we directly model the aggregated wind power produced by multiple turbines (e.g., all turbines in the wind farm). This approach lacks the flexibility of the microscopic approach, however it results in a simpler model as there is only a single quantity that is modeled (the aggregated power output).

The HMMs used in this study assume a discrete (in particular finite) state space. Though wind power output is a continuous variable, its measurements are rounded. To this end, we define N discrete levels (or states) to denote the possible values of the power output. For simplicity and in order to reduce the state space, we reduce the number of levels by mapping a set of data values to a level (e.g., by assigning a measurement to its nearest discrete level). We investigate the effect of this discretization on the probability distribution of the aggregated power output (by varying N), and in particular on the upper quantiles of the distribution. The "visible" part of the state-space of the HMM is either vector-valued, keeping track of the output of each wind turbine (in the microscopic approach), or single-valued (in the macroscopic approach). In both cases, we allow for an additional "hidden" component,

whose dynamics cannot be directly observed from the wind power output data.

The interpretation of the hidden process of an HMM (e.g., in physical terms) depends on the application. We hypothesize, for the case under consideration, that the hidden process can include meteorological conditions to which the wind turbines are subjected (see [25], [26] for an example of HMMs used in meteorology). These conditions are highly dependent in both space and time, resulting in dependencies between power output of turbines in close geographic proximity (e.g., turbines in the same wind farm) and high auto-correlation. To the best of our knowledge, modeling multiple output processes with a single hidden process has not been used before. In [27], one of the implemented approaches uses the Viterbi algorithm to estimate the parameters of a simple HMM from pseudo-measurements (with deterministic wind power generation and load patterns) and based on simulations the authors show that this approach increases the prediction accuracy, and that HMMs are more accurate than DTMCs.

The paper is organized as follows. In Section II, we discuss the post processing of the original data and the inherent dependency structure between the wind power outputs generated by the various turbines in the farm. In Section III, we describe the discretization of the measurement data. In Section IV, we present the results obtained using the HMM model for the microscopic approach and we compare them to the corresponding empirical results (obtained directly from the measurement data). In Section V, we present the results of the macroscopic approach. We present some conclusions in Section VI.

II. Description of the data

We have data from 6 wind turbines, (numbered 1 to 6) from an on-shore wind farm in the Netherlands covering the period of April 1 to September 12, 2016. The wind turbines are of the type ETW Directwind 900/52. The data is obtained through the Supervisory Control and Data Acquisition (SCADA) system of the wind turbines. For this wind farm, the measurements are stored as 20-minute averages.

The data include several variables in addition to the wind power (e.g., wind speed, wind direction, etc), however, for the purpose of this study, we only use the wind power data. Furthermore, since we are interested in the aggregated power output from all the turbines, we use only complete cases, i.e. only the time instances for which we have

data from all six wind turbines. Let P_t^w be the t-th (time-ordered) datum of the power output of turbine w, $w \in \mathcal{W} = \{1, \ldots, W\}$, $W = 6$, and let $\boldsymbol{P}^w = [P_1^w, \ldots, P_T^w]$ denote the complete time series data for turbine w, with $T = 8673$. Note that T represents the number of complete cases.

A. Analysis of complete data

It is expected that geographic proximity of the turbines leads to dependency in the data. This is validated in Table I, in which we present the correlation coefficients for the power output between turbine 6 and the other turbines. This is also validated in the produced scatter plots (see Section I of supplementary material).

TABLE I: Correlation between P_t^6 (power of turbine 6) and P_t^w, $w = \{1, \ldots, 5\}$ (power of turbines 1-5).

w	1	2	3	4	5
Correlation coefficient	0.886	0.909	0.929	0.916	0.954

III. discretization of data

We denote by $G_t = \sum_{w=1}^{6} P_t^w$ the total (aggregated) power produced by the six turbines together at time t, $t = 1, \ldots, T$. Let $\boldsymbol{G} = [G_1, \ldots, G_T]$ be the complete time series of G_t, $t = 1, \ldots, T$. The values of $\boldsymbol{G}$ lie in the interval $[-26.3, 5342.89]$ kW. Note that the negative values of $\boldsymbol{G}$ correspond to consumption of power by the wind turbines for performing internal functions like blade-pitch control, sensors, hydraulic breaks, etc.

In Section III-A, we discuss the discretization of the support of $\boldsymbol{P}^w$. In order to differentiate the discretized variables (with the binned support) from the original ones, all notation will be adapted to have a tilde, e.g., we denote by $\tilde{P}_t^w$ the discretized t-th measurement and with P_t^w the original t-th measurement of the power output of the w-th turbine. In the next section, we sketch the discretization procedure, and, in Section III-B, we compare the distributions as well as the $90\%, 95\%$ and 99% quantiles of the original and the discretized data.

A. Equidistant power levels and thresholds

For the discretization of the $\boldsymbol{P}^w$ variable, we define N levels, whose values are denoted by B_k, $k = \{1, \ldots, N\}$. In order to produce these values, we first consider the entire power support $[0, P_R]$, where P_R is the rated power of the turbines (the
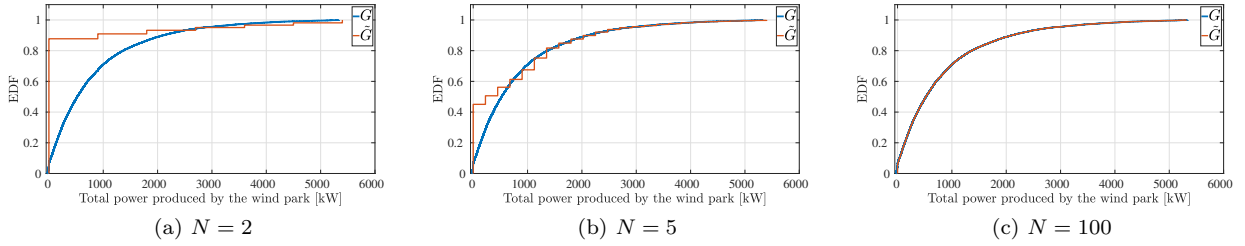
Fig. 1: Comparing EDFs of the total power $\boldsymbol{G}$ and its discretized counterpart $\tilde{\boldsymbol{G}}$, for different numbers of levels N.

theoretical maximum power a turbine of this type can produce). For the wind farm under consideration $P_R = 900$ kW. Secondly, we equidistantly divide $[0, P_R]$ in N levels. More concretely, this procedure produces $N-1$ thresholds, say $B_k^{\text{thres}} = \frac{B_k + B_{k+1}}{2}$, with $B_k = \left[\frac{k-1}{N-1}\right] P_R$, $k = \{1, \ldots, N\}$. Based on the above procedure, the data are discretized according to

$$\tilde{P}_t^w = \begin{cases} B_1, & \text{if } P_t^w < B_1^{\text{thres}}, \\ B_{k+1}, & \text{if } B_k^{\text{thres}} \leq P_t^w \leq B_{k+1}^{\text{thres}}, \\ & \quad \text{for } k = 1, \ldots, N-2, \quad \text{and} \\ B_N, & \text{if } P_t^w > B_{N-1}^{\text{thres}}, \end{cases} \tag{1}$$

for $t = 1, \ldots, T$ and $w \in \mathcal{W}$. We use the same $\{B_k\}$ and $\{B_k^{\text{thres}}\}$ values for all turbines. We would like to note that depending on the modeling objective (here it is mainly to capture the tail distribution of the wind power output), one could consider other approaches for the discretization of the data besides equal width, see [28] for a recent survey in discretization techniques.

B. Comparing measurement power with discretized measurement power

In this section, we analyze the effect of the discretization on the total power output distribution. We do so by comparing the distribution and the $90\%, 95\%$ and 99% quantiles of the original and discretized data ($\boldsymbol{G}$ and $\tilde{\boldsymbol{G}}$, respectively) varying the number of the levels, N.

1) Empirical distribution function (EDF) comparison: For a vector of data $\boldsymbol{Z} = [Z_1, \ldots, Z_T]$, the EDF is defined as $F(z) = \frac{1}{T} \sum_{t=1}^{T} \mathbb{1}(Z_t \leq z)$. The EDFs of $\boldsymbol{G}$ and $\tilde{\boldsymbol{G}}$, for $N = 2, 5, 100$, are depicted in Fig. 1a–1c. Note that, for $N = 5$ (Fig. 1b), the right tails match well. It should be noted that for $N = 5$ the EDFs do not match for smaller values of the support.

However for large N, the EDFs match over the entire support (see Fig. 1c).

2) Quantile comparison: To compare the $90\%, 95\%$ and 99% quantiles of $\boldsymbol{G}$ and $\tilde{\boldsymbol{G}}$, we first compute the corresponding threshold values (denoted G^*) for the $\boldsymbol{G}$ data. We define

$$\gamma_P := \frac{1}{T} \sum_{t=1}^{T} \mathbb{1}(G_t > G^*), \tag{2}$$

where $\mathbb{1}(\cdot)$ is the indicator function, and compute the value of G^* given γ_P. In particular, $\gamma_P = 0.1$ (90% quantile) yields $G^* = 2210.6$ kW, $\gamma_P = 0.05$ (95% quantile) yields $G^* = 2819.8$ kW, and $\gamma_P = 0.01$ (99% quantile) yields $G^* = 4411.3$.

Repeating the same analysis for the discretized data $\tilde{\boldsymbol{G}}$ for an increasing number of levels, N, produces as expected that $\gamma_P = \lim_{N \to \infty} \gamma_{\tilde{P}}^{(N)}$, with $\gamma_{\tilde{P}}^{(N)} = \frac{1}{T} \sum_{t=1}^{T} \mathbb{1}(\tilde{G}_t > G^*)$. This is illustrated in Fig. 2. Furthermore, we find that for $N = 5$ the relative error $RE = \frac{|\gamma_P - \gamma_{\tilde{P}}^{(N)}|}{\gamma_P} \times 100$ (in $\%$) equals $1.3\%, 0.69\%$ and 11%, for the $90\%, 95\%$ and 99% quantiles, respectively. Hence, for the study of the tail distribution (represented by the $90\%, 95\%$ and 99% quantiles) it is sufficient to choose $N = 5$, however, if the objective is to fit the entire distribution N should be chosen larger or the discretization approach implemented in Section III-A should be appropriately changed.

IV. Hidden Markov Models for the microscopic approach

In this section, we consider HMMs (for an introductory description of HMMs see Section II of the supplementary material) and [29]. Under the HMMs formalism, the individual wind turbine power output is discrete and Markovian in nature, but it also depends on the evolution of the hidden component of
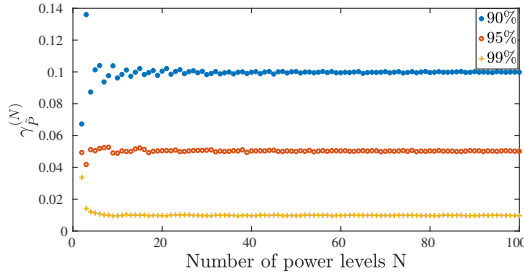
Fig. 2: Comparing $\gamma_{\tilde{P}}^{(N)}$ for different quantile thresholds, G^*, as a function of the number the levels, N.

the Markov process. HMMs offer a flexible framework that can integrate different layers of complexity (by accounting for the joint state description of all the turbines in the park) and of information for the spatio-temporal dynamics (environmental and geographic conditions) captured in the hidden process, surpassing the capabilities of copula modeling [19].

A. Individual stochastic modeling (matrix) per turbine

Let $\{Y_t^w, t = 1, 2, \ldots\}$, $w \in \mathcal{W}$ denote the discretized wind power output of the individual wind turbines at time t, with state-space $\mathcal{N} = \{B_1, \ldots, B_N\}$. In the HMMs framework, we assume that the stochastic processes $\{Y_t^w\}$ depend on a sequence of hidden states, $\{X_t, t = 1, 2, \ldots\}$, with state space $\mathcal{M} = \{1, \ldots, M\}$. More specifically, the processes $\{Y_t^w\}$ are governed by the stochastic matrices $L^{w,i} = \left(L_{\mu,\nu}^{w,i}\right)_{\mu,\nu \in \mathcal{N}}$ whenever $\{X_t = i\}$, and the transition probabilities of $\{X_t, t = 1, 2, \ldots\}$ are given by the stochastic matrix $A = (A_{i,j})_{i,j \in \mathcal{M}}$, i.e.,

$$L_{\mu,\nu}^{w,i} = \mathbb{P}\left(Y_{t+1}^w = \nu | Y_t^w = \mu, X_{t+1} = i\right)$$
$$A_{i,j} = \mathbb{P}\left(X_{t+1} = j | X_t = i\right). \tag{3}$$

For $t = 1$, the initial distribution of X_1 is given by $\pi_i = \mathbb{P}(X_1 = i)$, and of Y_1^w by $\rho_\mu^{w,i} = \mathbb{P}(Y_1^w = \mu | X_1 = i)$. Note that $M = 1$ implies that there is no hidden state and the model simplifies to a simple DTMC, see [30].

Our goal is to estimate from the discretized data, $\tilde{P}^w$, the parameters of the above mentioned model: $\lambda = (\pi, \rho, A, L)$, with $\pi = (\pi_i)_{i \in \mathcal{M}}$, $A = (A_{ij})_{i,j \in \mathcal{M}}$, $\rho = \left(\rho_\mu^{w,i}\right)_{\mu \in \mathcal{N}}^{w \in \mathcal{W}, i \in \mathcal{M}}$ and $L = \left(L_{\mu\nu}^{w,i}\right)_{\mu,\nu \in \mathcal{N}}^{w \in \mathcal{W}, i \in \mathcal{M}}$. For the estimation of the model parameters, we use the Expectation-Maximization (EM) algorithm for HMMs, see [31]–[37]. A detailed

explanation of the EM algorithm and the model parameter estimation is provided in Section III of the supplementary material. Also see Section V of the supplementary material for complexity evaluation and improvements of HMMs.

1) Quantile comparison: Given the discretized data $(\tilde{P}^w)$ and given the number of hidden states (M), we estimate the parameters of the HMM (see Section III of the supplementary material). We generate surrogate wind power output measurements, denoted by $\hat{P}_{\text{HMM}}^w$, based on the HMM, and let $\hat{G}_t^{\text{HMM}} := \sum_{w=1}^{6} \hat{P}_{t,\text{HMM}}^w$ be the total power surrogate time series. We calculate the fraction of time $\hat{G}_t^{\text{HMM}}$ is greater than the quantile thresholds, G^* (see section III-B),

$$\gamma_{\hat{P}_{\text{HMM}}}^{(N,M)} = \frac{1}{T} \sum_{t=1}^{T} \mathbb{1}(\hat{G}_t^{\text{HMM}} > G^*). \tag{4}$$

The numerical procedure can be summarized as follows: We first estimate the HMM parameters, thereafter we generate 100 independent realizations of $\hat{P}_{t,\text{HMM}}^w$, $t = 1, \ldots, T$, $T = 10^5$, and we compute the fraction $\gamma_{\hat{P}_{\text{HMM}}}^{(N,M)}$ for each of these realizations. The mean and standard deviation of these 100 fractions are denoted by $\bar{\gamma}_{\hat{P}_{\text{HMM}}}^{(N,M)}$ and $\sigma(\gamma_{\hat{P}_{\text{HMM}}}^{(N,M)})$, respectively. In order to comment on the model best fitting the given data set, we compare the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) statistic values. The AIC and the BIC values, for a given model, are computed as follows

$$\text{AIC} = 2p - 2\log\widehat{\mathcal{L}},$$
$$\text{BIC} = p\log T - 2\log\widehat{\mathcal{L}}, \tag{5}$$

with p the number of free parameters in the model, $\widehat{\mathcal{L}}$ the maximized value of the likelihood function for the parameter values $\widehat{\lambda}$ for the data set, and T the number of observations in the data set, see [39].

We compared $\bar{\gamma}_{\hat{P}_{\text{HMM}}}^{(N,M)}$ for different quantile thresholds, G^*, with γ_P for $N = 5$ and different number of hidden states M. We find that the relative error of $\bar{\gamma}_{\hat{P}_{\text{HMM}}}^{(N,M)}$ and γ_P, $\text{RE} = \frac{|\gamma_P - \bar{\gamma}_{\hat{P}_{\text{HMM}}}^{(N,M)}|}{\gamma_P} \times 100$ (in %) achieves its minimum (for all three quantiles) for $M = 9$. We also find that the lowest value of $-\log\widehat{\mathcal{L}}$ and the AIC is achieved for $M = 9$. However, the BIC value grows with M as the penalty term for the BIC is higher than in the AIC. In Table II, we show the results for $N = 5, M = 9$. For more results and comparisons on different values of M, the interested reader is referred to Table I of the supplementary material.

TABLE II: Results on the microscopic approach with individual stochastic matrices (cf. Section IV-A) for $N = 5$.

M	Quantile	$\bar{\gamma}^{(5,M)}_{\hat{P}_{\text{HMM}}} \pm \sigma(\gamma^{(5,M)}_{\hat{P}_{\text{HMM}}})$	RE (%)	$-\log\widehat{\mathcal{L}} \times 10^{-4}$	AIC $\times 10^{-4}$	BIC $\times 10^{-4}$	p
	90%	0.1011 ± 0.0032	1.1512				
9	95%	0.0517 ± 0.0020	3.2605	2.4011	5.0776	6.0500	1376
	99%	0.0114 ± 0.0060	13.127				

By comparing the quantile values, the $-\log\widehat{\mathcal{L}}$, and the AIC test, we can conclude that for $N = 5$, the optimal number of hidden states is $M = 9$. For further experimentation, we also compare the EDFs of the total power from the original data $\boldsymbol{G}$, the discretized data, $\tilde{\boldsymbol{G}}$, and the HMMs time series, $\hat{\boldsymbol{G}}^{\text{HMM}}$, for $N = 5$ and $M = 9$. As expected, the HMM is able to reproduce the right tail of the EDF. For more results and comparisons on different values of M, the interested reader is referred to Fig. 3 of the supplementary material). We would like to note that if the objective were to fit the entire distribution without drastically increasing the values of N and M, then we would propose to use a moment matching based fitting method, see, e.g., [38] and the references therein.

B. Same stochastic modeling (matrix) per turbine

In (3), we define separate stochastic matrices for each wind turbine (indexed by w). For this model setup, the number of parameters is $p = M^2 + MWN^2 - MN - 1$. For example, for $N = 5, M = 9$ and $W = 6$, the number of parameters to be estimated is $p = 1376$. In order to reduce the number of estimated parameters, in this section, we investigate a model setup in which the same (set of) stochastic matrices is used for each of the wind turbines. Thus, the transition probability matrix of each power output process $\{Y_t^w\}$ depends on the hidden state, but not on the turbine index w, i.e., the matrix $L^{w,i}$ in (3) is now independent of w. This reduces the number of parameters of the model. The details of the parameter estimation are given in Section VI of the supplementary material. We study the trade-off between reducing the number of parameters of the model by calculating anew the $-\log\widehat{\mathcal{L}}$, AIC and BIC values, and by comparing them with the values reported in Table II.

The number of parameters to be estimated for this model are $p = M^2 + MN^2 - M - 1$. We follow the same procedure as in Section IV-A1 for the estimation of the quantile fractions, denoted by $\gamma'^{(N,M)}_{\hat{P}_{\text{HMM}}}$ for this model. Similarly, to the previous section, the best

results are obtained for $N = 5$, $M = 9$. In Table III, we present the results for $N = 5$, $M = 9$, for the case of the same stochastic matrix for all the turbines. For more results and comparisons on the different values of M, the interested reader is referred to Table II of supplementary material.

Comparing the values of $-\log\widehat{\mathcal{L}}$, AIC and BIC in Table II and Table III, it is evident that $-\log\widehat{\mathcal{L}}$ is slightly lower for the former. This implies that having different stochastic matrices (as in Section IV-A) for different wind turbines fits the given data better than having the same stochastic matrices (as in Section IV-B). Of course, one needs to also take into account the number of unknown parameters that need to be estimated. For the first approach these amount to 1376 parameters. For the second approach there are 296 parameters in case $N = 5, M = 9$. Furthermore, we find that the AIC and the BIC values are higher for the first approach, but the difference is small.

V. Results of the Macroscopic approach

In the microscopic approach studied in Section IV, we modeled the individual wind turbine power output with HMMs, with the aim to model the high quantiles of the distribution of the total power produced by all turbines together ($\boldsymbol{G}$). As discussed in the introduction, a different approach is to model $\boldsymbol{G}$ directly. In this section, the measurement data of $\boldsymbol{G}$ are discretized (cf. Section III), and modeled with a HMM (similar to the procedure of Section IV).

We recall from Section III-B2 that the thresholds for the 90%, 95% and 99% quantiles of $\boldsymbol{G}$ are 2210.6, 2819.8 and 4411.3 kW, respectively. Discretizing $\boldsymbol{G}$ directly, we find that with 15 levels ($N = 15$) the relative errors are 0.92%, 5.8% and 4.6%, for the 90%, 95% and 99% quantiles, respectively. One could further reduce the relative error by increasing N, but we consider this to be out of the scope of the paper, and for this reason restrict our analysis to $N = 15$, as this seems to adequately capture the high quantiles of the distribution of $\boldsymbol{G}$.

TABLE III: Results on the microscopic approach with the same stochastic matrices (cf. Section IV-B) for $N = 5$.

M	Quantile	$\bar{\gamma}'^{(5,M)}_{\tilde{P}_{\text{HMM}}} \pm \sigma(\gamma'^{(5,M)}_{\tilde{P}_{\text{HMM}}})$	RE (%)	$-\log\widehat{\mathcal{L}} \times 10^{-4}$	AIC $\times 10^{-4}$	BIC $\times 10^{-4}$	p
9	90%	0.0987 ± 0.0032	1.243	2.4724	5.0040	5.2132	296
	95%	0.0517 ± 0.0020	3.391				
	99%	0.0108 ± 0.0060	8.116				

A. Quantile fraction comparison

Similar to the analysis of the previous sections, we generate 100 realizations, and compare the quantile values $\gamma^{(N,M)}_{\bar{G}}$, and their respective relative errors while varying the number of hidden states M. Furthermore, we compute the corresponding $-\log\widehat{\mathcal{L}}$, AIC and BIC values. Based on this procedure, we find that with $M = 1$, the quantile fractions are very well captured, however the values of $-\log\widehat{\mathcal{L}}$, AIC and BIC are much higher compared to the $M \geq 2$ cases. The lowest value of $-\log\widehat{\mathcal{L}}$ is obtained with $M = 5$. Note that the AIC and the BIC values increase in M (for $M \geq 2$) and there does not seem to exist a minimum in the AIC value, in contrast to the microscopic approach. We depict the results for $M = 5$ in Table IV. For more results and comparisons for different M values see Table III of the supplementary material.

We point out that in this case the number of parameters ($p = 1144$) exceeds the number in the microscopic approach for the same matrix ($p = 296$), see Section IV-B. This is related to the higher number of levels (N) needed in the macroscopic approach. In the microscopic approach, with W turbines and N levels for each turbine output, the total (i.e., aggregated) discretized power output has WN possible states (e.g., in the case under consideration $W = 6$ and $N = 5$). By contrast, in the macroscopic approach the discretized total output has N states, therefore N must be larger so as to capture the tail of the distribution with the level of accuracy.

Furthermore, we note that one cannot directly compare the $-\log\widehat{\mathcal{L}}$, AIC and BIC values of the model in Table IV with the models in Table II and Table III as the former uses the observation data set $\hat{G}$ (directly discretizing G) and the later two use $\tilde{P}^w$ (obtained from discretizing the individual wind turbine power output).

VI. Conclusions

In this paper, we studied the performance of HMMs for modeling the power output of a wind farm. Such models have a rather simple structure, making them suitable for theoretical analysis. Despite their simplicity, they are able to accurately reproduce the distribution of the wind farm power output, as demonstrated in this study. In particular, these models are well able to capture the tail of the distribution.

We considered two approaches: in the microscopic approach, the power output of individual turbines were modeled with Markov chains dependent on a single hidden process, and subsequently aggregated to obtain the total power output. With the HMM we were able to account for the dependence structure among the wind turbines, resulting in a precise reproduction of the upper quantiles. With the macroscopic approach, we directly modeled the aggregated output of all turbines.

To assess the quality of the estimated models, we computed their log-likelihoods, as well as the Akaike and Bayesian information criteria (AIC and BIC). With AIC, one can take into account the number of model parameters, whereas the BIC accounts for both the number of parameters and the amount of data used for estimation. In the microscopic approach, the HMMs in which all turbines had the same set of stochastic matrices (discussed in section IV-B) performed most favorable in terms of AIC and BIC. This is due to the smaller number of parameters needed in these models, compared to the HMMs with different sets of stochastic matrices for the different turbines (section IV-A). The HMMs from the macroscopic approach also required more parameters than the HMM from the microscopic approach with the same set of stochastic matrices (as discussed in section V).

In this study we have shown that simple discrete HMMs are able to reproduce the distribution of wind farm power output, in particular its upper quantiles. Due to their simple structure, these models are amenable to theoretical analysis using methods from e.g., the field of queuing theory. We will report on such analysis in a follow-up study.

TABLE IV: Results on the macroscopic approach for $N = 15, M = 5$

M	Quantile	$\bar{\gamma}_{\widehat{G}}^{(15,M)} \pm \sigma(\gamma_{\widehat{G}}^{(15,M)})$	RE (%)	$-\log\widehat{\mathcal{L}} \times 10^{-4}$	AIC $\times 10^{-5}$	BIC $\times 10^{-5}$	p
5	90%	0.0988 ± 0.0102	1.154				
	95%	0.0475 ± 0.0057	5.095	0.9761	0.2181	0.2989	1144
	99%	0.0101 ± 0.0022	0.507				

Acknowledgment

References

[1] R. Karki and R. Billinton, "Reliability/cost implications of PV and wind energy utilization in small isolated power systems," IEEE Transactions on Energy Conversion, vol. 16, no. 4, pp. 368–373, 2001.

[2] P. Hu, R. Karki, and R. Billinton, "Reliability evaluation of generating systems containing wind power and energy storage," IET Generation, Transmission & Distribution, vol. 3, no. 8, pp. 783–791, 2009.

[3] A. S. Dobakhshari and M. Fotuhi-Firuzabad, "A reliability model of large wind farms for power system adequacy studies," IEEE Transactions on Energy Conversion, vol. 24, no. 3, pp. 792–801, 2009.

[4] S. Klein and W. Beckman, "Loss-of-load probabilities for stand-alone photovoltaic systems," Solar Energy, vol. 39, no. 6, pp. 499–512, 1987.

[5] M. Shafiee and J. D. Sørensen, "Maintenance Optimization and Inspection Planning of Wind Energy Assets: Models, Methods and Strategies", Reliability Engineering & System Safety, In Press, 2017.

[6] N. B. Negra, O. Holmstrøm, B. Bak-Jensen, and P. Sørensen, "Model of a synthetic wind speed time series generator," Wind Energy, vol. 11, no. 2, pp. 193–209, 2008.

[7] A. Simonian and J. Guibert, "Large deviations approximation for fluid queues fed by a large number of on/off sources," IEEE Journal on Selected Areas in Communications, vol. 13, no. 6, pp. 1017–1027, 1995.

[8] P. Giorsetto and K. F. Utsurogi, "Development of a new procedure for reliability modeling of wind turbine generators," IEEE Transactions on Power Apparatus and Systems, no. 1, pp. 134–143, 1983.

[9] R. Karki, P. Hu, and R. Billinton, "A simplified wind power generation model for reliability evaluation," IEEE Transactions on Energy Conversion, vol. 21, no. 2, pp. 533–540, 2006.

[10] G. Sansavini, R. Piccinelli, L. Golea, and E. Zio, "A stochastic framework for uncertainty analysis in electric power transmission systems with wind generation," Renewable Energy, vol. 64, pp. 71–81, 2014.

[11] J. Zeng and W. Qiao, "Short-term wind power prediction using a wavelet support vector machine," IEEE Transactions on Sustainable Energy, vol. 3, no. 2, pp. 255–264, 2012.

[12] A. Feijoo and D. Villanueva. "Wind farm power distribution function considering wake effects". IEEE Transactions on Power Systems, vol. 32, no. 4, pp. 3313-3314, 2017.

[13] A. Feijoo and D. Villanueva. "Contributions to wind farm power estimation considering wind direction?dependent wake effects." Wind Energy, vol. 20, no. 2, pp. 221-231, 2017.

[14] S. H. Hosseini, C. Y. Tang, and J. N. Jiang, "Calibration of a wind farm wind speed model with incomplete wind data," IEEE Transactions on Sustainable Energy, vol. 5, no. 1, pp. 343–350, 2014.

[15] S. Kolumbán, S. Kapodistria, and N. Nooraee. "Short and long-term wind turbine power output prediction." arXiv preprint arXiv:1707.06497, 2017.

[16] P. Chen, P. Siano, B. Bak-Jensen, and Z. Chen, "Stochastic optimization of wind turbine power factor using stochastic model of wind power," IEEE Transactions on Sustainable Energy, vol. 1, no. 1, pp. 19–29, 2010.

[17] G. Papaefthymiou and B. Klockl, "MCMC for wind power simulation," IEEE Transactions on Energy Conversion, vol. 23, no. 1, pp. 234–240, 2008.

[18] F. C. Sayas and R. Allan, "Generation availability assessment of wind farms," IEE Proceedings-Generation, Transmission and Distribution, vol. 143, no. 5, pp. 507–518, 1996.

[19] G. Papaefthymiou and D. Kurowicka, "Using copulas for modeling stochastic dependence in power system uncertainty analysis," IEEE Transactions on Power Systems, vol. 24, no. 1, pp. 40–49, 2009.

[20] P. Li, X. Guan, J. Wu, and X. Zhou, "Modeling dynamic spatial correlations of geographically distributed wind farms and constructing ellipsoidal uncertainty sets for optimization-based generation scheduling," IEEE Transactions on Sustainable Energy, vol. 6, no. 4, pp. 1594–1605, 2015.

[21] D. D. Le, G. Gross, and A. Berizzi, "Probabilistic modeling of multisite wind farm production for scenario-based applications," IEEE Transactions on Sustainable Energy, vol. 6, no. 3, pp. 748–758, 2015.

[22] M. Khodayar, O. Kaynak, and M.E. Khodayar. "Rough Deep Neural Architecture for Short-term Wind Speed Forecasting." IEEE Transactions on Industrial Informatics, DOI 10.1109/TII.2017.2730846, 2017.

[23] A.S. Qureshi, A. Khan, A. Zameer, and A. Usman. "Wind Power Prediction using Deep Neural Network based Meta Regression and Transfer Learning." Applied Soft Computing, vol. 58, pp. 742–755, 2017.

[24] W. S. Wadman, D. T. Crommelin, and B. P. Zwart, "A large-deviation-based splitting estimation of power flow reliability," ACM Transactions on Modeling and Computer Simulation (TOMACS), vol. 26, no. 4, p. 23, 2016.

[25] F.O. Hocaoğlu, O.N. Gerek, and M. Kurban. "A novel wind speed modeling approach using atmospheric pressure observations and hidden Markov models." Journal of Wind Engineering and Industrial Aerodynamics, vol. 98, no. 8, pp.472-481, 2010.

[26] C. Franzke, D. Crommelin, A. Fischer, and A. J. Majda, "A hidden markov model perspective on regimes and metastability in atmospheric flows," Journal of Climate, vol. 21, no. 8, pp. 1740–1757, 2008.

[27] S. Jafarzadeh, S. Fadali, C.Y. Evrenosoglu, and H. Livani. "Hour-ahead wind power prediction for power systems using hidden Markov models and Viterbi algorithm." IEEE Power and Energy Society General Meeting, pp. 1-6, 2010.

[28] S. Kotsiantis and D. Kanellopoulos. "Discretization techniques: A recent survey." GESTS International Transactions on Computer Science and Engineering, vol. 32, no. 1, pp. 47–58, 2006.

[29] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," Proceedings of the IEEE, vol. 77, no. 2, pp. 257–286, 1989.

[30] J. R. Norris, Markov Chains. Cambridge University Press, 1998, no. 2.

[31] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," Journal of the Royal Statistical Society. Series B (Methodological), pp. 1–38, 1977.

[32] I. Horenko, E. Dittmer, A. Fischer, and C. Schütte, "Automated model reduction for complex systems exhibiting metastability," Multiscale Modeling & Simulation, vol. 5, no. 3, pp. 802–827, 2006.

[33] Z. Ghahramani and M. I. Jordan, "Supervised learning from incomplete data via an EM approach," Advances in Neural Information Processing Systems, pp. 120–120, 1994.

[34] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains," The Annals of Mathematical Statistics, vol. 41, no. 1, pp. 164–171, 1970.

[35] J. A. Bilmes, "A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models," International Computer Science Institute, vol. 4, no. 510, p. 126, 1998.

[36] M. I. Jordan and L. Xu, "Convergence results for the EM approach to mixtures of experts architectures," Neural Networks, vol. 8, no. 9, pp. 1409–1431, 1995.

[37] C. J. Wu, "On the convergence properties of the EM algorithm," The Annals of Statistics, pp. 95–103, 1983.

[38] G. Horváth and M. Telek, "Fitting methods based on distance measures of marked Markov arrival processes". In Seminal Contributions to modeling and Simulation, Springer International Publishing, pp. 159–183, 2016.

[39] K. P. Burnham and D. R. Anderson, "Multimodel inference understanding AIC and BIC in model selection," Sociological Methods & Research, vol. 33, no. 2, pp. 261–304, 2004.

Stella Kapodistria (B.Sc. Mathematics 2004, M.Sc. Statistics and Operations Research 2007, Ph.D. Mathematics 2009) is an Assistant Professor at the Department of Mathematics and Computer Science at Eindhoven University of Technology (TU/e), since 2014. She serves on the editorial board of the Probability in the Engineering and Informational Sciences international journal and is a guest editor of the Annals of Operation Research international journal. Since 2013, she has been involved in the development of prognostic and diagnostic algorithms for the DAISY and the DAISY4Offshore projects on wind turbines.

Bert Zwart (M.A. Econometrics 1997, Ph.D. Mathematics 2001) is leader of the CWI Stochastics group (Amsterdam) and Professor at the Eindhoven University of Technology). Recent focus of his work is on the applications of probabilistic methods to power systems. He is co-organizer of a special semester on the Mathematics of energy systems taking place in Cambridge UK, spring 2019.

Debarati Bhaumik received her B.Sc. degree in Physics from University of Delhi, India in 2010 and the M.Sc. degree in Physics from Indian Institute of Technology Madras, India in 2012. She is a Ph.D. student in the Scientific Computing group at CWI Amsterdam. Her Ph.D. topic is on reliability assessment of power grids integrated with wind generations and energy storage. Her present research interests include stochastic modeling, integration of renewables in power network, and Monte Carlo techniques for rare-event simulation.

Daan Crommelin received his Ph.D. from Utrecht University in 2003 and worked as a postdoc at New York University's Courant Institute, before joining CWI Amsterdam in 2006. Currently he is leader of the Scientific Computing research group at CWI and a Professor at the Korteweg-de Vries Institute for Mathematics of the University of Amsterdam. His research interests include stochastic and computational methods for multi-scale dynamical systems, uncertainty quantification, rare event simulation, statistical inference, applications in atmosphere-ocean-climate science and energy networks.